Please, note that this is an Accepted Manuscript of an article published by Taylor & Francis Group in European Planning Studies on 2014 available online: http://www.tandfonline.com/doi/full/10.1080/09654313.2012.752439#.U5gKrJTV-A8.

Regional scientific production and specialization in Europe

The role of HERD

MANUEL ACOSTA

DANIEL CORONADO

ESTHER FERRÁNDIZ

M. DOLORES LEÓN

Departamento de Economía General

University of Cadiz

(Spain)

Correspondence address: Departamento de Economía General. Facultad de Ciencias Económicas y Empresariales, Universidad de Cádiz, Glorieta Carlos Cano s/n, 11002 Cadiz, Spain. Email: daniel.coronado@uca.es

Abstract: This paper analyses the effects of R&D expenditure in the higher education sector (HERD) on the scientific production across regions in Europe 15. Our research questions relates to the regional production of science and the role of academic R&D expenditures on regional scientific output. The results show that money affects the production of scientific results in regions. On average, we found different impacts and lags of R&D expenditure according to the level of regional development. Our findings also suggest that scientific specialization is a significant factor affecting scientific outputs, although its effects differ across disciplines and regions.

Key words: Scientific production, R&D expenditure in the higher education sector, Europe-15 regions, regional scientific specialization.

JEL codes: O18, O31, O32.

1. Introduction

The identification of academic scientific capacities is important for regional and supranational governments that must decide scientific priorities and the allocation of funds. This issue is particularly relevant in Europe, where there is an intense debate about how the current situation and trends in research could have a negative influence on competitiveness and employment in the years ahead (European Commission, 2000, 2007). Moreover, from a political viewpoint, the publication of the Commission's paper on the European Research Area (ERA) in 2000 has stressed the importance of regions in the development of research and innovation capacity within Europe.

The European Commission specifically identifies a role for the regional geographical classification in achieving an ERA (European Commission, 2001) and implementing the Europe 2020 Strategy (European Commission, 2010). Among others, funding mechanisms are considered relevant instruments in shaping the quantity and quality of research (de Dominicis et al., 2011). However, while there is some evidence of the short-term usefulness of research and development (R&D) incentives at the country level to promote scientific research (e.g. Adams and Griliches, 1998; Crespi and Geuna, 2008), regional-specific information seems to be largely missing in the literature. In this paper, we fill this gap by providing some insight into scientific production across European regions. Additionally, we particularly focus on the role of funding in the development of university-based research at the regional level. The methodology involves a descriptive analysis and several econometric models to estimate the impact of university funds in encouraging the production of science in European regions.

This paper contributes to the literature in several ways. First, despite the relevance of some economic aspects of research activities in universities (see the surveys by Dasgupta and David, 1994; Stephan, 1996), the empirical literature concerning the production of science in universities at the regional level is scarce and this paper provides new evidence on this topic. Second, we draw on a new data set to address our research question by using regionalized academic published papers retrieved from the Thomson Reuters Web of Science (WoS). The data contain about one million papers published in the period 1998–2004 classified by regions in Europe 15 (NUTS II level of aggregation). Third, from a political viewpoint, it provides policymakers with a direct contribution to mapping science, some insights on the role of academic R&D funds, and thus some clues for a better knowledge of the ERA and viable regional specialization opportunities.

This paper is organized as follows. Section 2 summarizes the literature relevant to this paper. Section 3 describes the data and provides an overview of the patterns of university scientific production at the regional level across Europe 15. Section 4 presents a regional version of a knowledge production function (KPF) and provides estimates of several models explaining the effects of R&D expenditures on the production of scientific knowledge. Section 5 summarizes the main findings and discusses the policy implications.

2. Theoretical and empirical background

This literature review is organized around two questions relevant to this paper: (i) Why is university research important for regional economics? (ii) What is the role of R&D expenditure in explaining the production of science?

The positive effects of universities in regions may occur through a variety of university outputs that potentially have important impacts on regional economic development. In this paper, we focus on one of these outputs: the production of scientific knowledge. University scientific knowledge may have an influence on innovation in regions in different ways. On the one hand, there is a potential direct contribution when a university produces useful new scientific knowledge with applications to industrial processes. The papers by Mansfield (1991, 1998), Mansfield and Lee (1996), Cohen et al. (1998), Beise and Stahl (1999), among others, have emphasized that knowing the characteristics of scientific production in universities and their underlying mechanisms is useful for their contribution to the development of industrial innovations. On the other hand, the production of university knowledge may have an indirect contribution to regional innovation because of the flow of knowledge between universities and firms. This knowledge interaction can take place through a variety of channels between academics and firms (when reading scientific papers, or via direct conversation or informal meetings with the inventors, etc.). The flow of knowledge has important potential benefits for regions because of spillovers from university to industry affecting not only technology, but other relevant variables for the economic system (Jaffe, 1989; Jaffe et al., 1993; Anselin et al., 1997; Anselin et al., 2000; Verspagen and Schoenmakers, 2000; Maurseth and Verspagen, 2002; Acosta and Coronado, 2003;

Agrawal and Cockburn, 2003; Fischer and Varga, 2003; Audretsch et al., 2005; Calderini and Scellato, 2005; Abramovsky et al., 2007; Acosta et al., 2011a). The proliferation of a consistent literature illustrating the importance of physical proximity for knowledge flows and for the promotion and development of innovation and new firm formation, along with the high degree of self-government enjoyed by many European regions, makes it clear that the study of university knowledge is relevant not only in national or supranational contexts, but also at the regional level.

This literature has addressed the production of knowledge in universities from the perspective of its consequences, and overall it stresses that the production of academic knowledge in general and scientific knowledge in particular is important for the economic system. We assume then, that the stronger the capacities for production of knowledge, the more beneficial their effect should be. However, little is known about the factors that can strengthen these scientific capacities, and particularly about the effects of the amount of R&D funds that universities receive. In the following paragraphs, we summarize the main results concerning the effects of funding on the production of science with a particular focus on the role of university R&D expenditure in promoting the production of science in universities.[1]

Two papers by Adams and Griliches (1996, 1998) were among the early attempts to measure the relation between inputs (R&D expenditures) and outputs (scientific publications and citations) from an economic viewpoint. Their point of departure was evidence of a discrepancy between the growth of R&D expenses (5.5% per year in real

---

[1] A stream of literature focused on the individual productivity of researchers has sometimes considered R&D funding as an "environmental attribute", along with other

terms) and the total number of scientific articles (1% per year) for the US during 1981–1991. Several regressions using different lags for R&D provided an average elasticity of 0.6 for papers and 0.7 for citations at the university and field level, suggesting the possibility of diminishing returns to scale. However, the results were possibly biased because, as the author remarked, spillover effects among universities and fields were not taken into account and a difference in elasticity estimates using more aggregated models is possible. Therefore, serious data limitations and difficulties hindered the authors from drawing firm conclusions.

In subsequent research, Adams et al. (2005) studied the size of scientific teams and institutional collaboration with data derived from 2.4 million scientific papers written within the 110 top US research universities that had at least one author from this set of leading US universities. Their analysis was carried out over the period 1981–1999. The source of their data was the WoS. They found positive and highly significant coefficients of the logarithm of the lagged stock of R&D (with values around 0.45 for the equation of log (papers) as the dependent variable, and 0.55 for the equation of log (citations) as the dependent variable), suggesting diminishing returns to the stock of R&D applied at the university-field level.

Following a similar methodology to Adams and Griliches (1998), Crespi and Geuna (2008) examined the determinants of scientific production at the cross-country level. Their data contain a sample of 14 countries and 22 years (1981–2002) for which the authors had information on Higher Education R&D (HERD) expenditures. The outputs (number of papers and citations) were taken from the Thomson Reuters national science indicators database on published papers and citations. This research differs from

Adams and Griliches (1998) in several respects (different structure of data, context, citation, etc.), but mainly in considering the spillover effects of HERD in the original KPF. While difficulties exist in obtaining robust results for elasticities of the outputs, given the poor quality of the data and modelling problems, their models suggested decreasing returns to the domestic component of R&D. The analysis of international spillovers indicated evidence of a significant impact from the weighted investment in HERD in other countries.

Payne and Siow (2003) estimated the effects of federal research funding on research outcomes in the US at the university level using publications as the measure of research outcomes from WoS data on articles published and citations to articles published. Their analysis of these outputs covered 57 universities and 1,017 observations, representing about 18 years of data for each university. For scientific articles as the dependent variable, all their estimations for federal research funding were significant and showed diminishing returns. They also used citations per article but obtained a negative and very small effect. The authors concluded that increasing federal research funding results in more, but not necessarily higher quality, research output.

In a study on European universities, Aghion et al. (2007) used a survey questionnaire sent to the European universities in the 2006 Top 500 Shanghai ranking. Using regression analysis, they found a significant and positive relationship between budget per student and research performance at the country level. Their analysis also indicates that the research performance of universities is positively associated with their size and their age.

The main lesson from this empirical literature is that money helps to achieve a better research performance. However, difficulties in obtaining accurate data prevent the reliable estimation of university R&D effects (elasticities), although most of the analyses found decreasing returns to university R&D expenditures. Moreover, as mentioned at the beginning of this paper, despite the role of regions in the research policies framed into the ERA strategy, there is no previous research on the effects of R&D funds on scientific production at the regional level.

In order to contribute to this empirical literature, we address a number of related empirical questions in the following sections:

1. How is the production of scientific research distributed across European regions? What regions lead the generation of science and in what fields? What are the regional specialization patterns across European regions? How are they performing in the production of science? Are they exploiting their scientific competitive advantages?

2. What are the effects of academic R&D funding in promoting the production of scientific research at the regional level? What are the time lags of these effects? Is there any difference according to the regional level of economic development?

3. University scientific production and specialization across European regions

In order to cope with the first group of questions, in this section we provide an overview of the distribution of university research for 1998–2004 in European Union 15 (EU-15) using an original data set. Some findings at the more detailed level of individual regions and fields over the same period are also presented.

3.1 Data description

The data set used in this study consists of a set of 1,206,644 university research articles published in scientific journals indexed by the Science Citation Index Expanded (SCI) in the period 1998–2004 and classified by European regions. The SCI is part of the WoS, which is a bibliographical database produced by Thomson Reuters. The main advantage of WoS is that it provides a complete list of all authors and their affiliations. There are also some well-known limitations of this database. For example, it does not include all journals, and the WoS journal list is biased strongly towards journals published in English (for details, see Bordons et al., 2002; van Raan, 2005; Weingart, 2005).

Our database was built as follows:

1. Data on academic publications containing at least one author affiliated with a university from an EU-15 country for 1998–2004 were retrieved from the SCI. It is worth noting that the lack of normalization in the way in which academic institutions are named hinders the finding of academic publications. For this reason, we included several search terms to help identify higher education institutions in both English and other languages (e.g. *fachhochschule*, *yliopisto*, *ecole*, *institut nacional polytehcnique*, *politécnico*, *scuola*, *hogskola*, *universitet*, etc.). This search resulted in 994,938 publications.

2. The second step involved regionalization at the NUTS II level of aggregation of the academic publications obtained in step 1 (213 regions). We first identified the NUTS II associated with each university using the list provided by the members of the European Indicators, Cyberspace and the Science–Technology–Economy System (EICSTES). For those universities not included in the EICSTES list, we searched for the address on each university's web site and matched it with the relevant region. Then, all publications were grouped by region. In the case of publications involving multiple regions, full counts were applied to all regions involved (i.e. crediting one publication to each region). As a result, in this step we obtained 1,206,644 publications. In this step, it is important to note the concern expressed by Hoekman et al. (2009) with respect to multiple affiliations. In most cases, the multiple regions involved in a single publication are associated with different researchers; however, a single researcher may have multiple affiliations (e.g. if he/she works for two or more universities) and then report more than one address in the publication. In these cases, the full-count method was also applied, crediting one publication to each listed region. Regarding the counting method, despite the fractional and full-count methods both being accepted widely and both having advantages and disadvantages (Okubo & Zitt, 2004), we used the full-count method for three main reasons.[2] First, we assumed that "each author, main institution and country listed in the affiliated addresses made a non-negligible contribution" and thus deserve full credit (Tijssen & van Leeuwen, 2003). Second, given that we deal with

[2] In a fractional count, the credit for a publication is divided equally among all authors. In contrast, in a full count, each author is credited with one publication. For a review and comparison of publication counts obtained using different methods, see Gauffriau et al. (2008).

a considerable volume of data, full counting becomes the most convenient solution (van Raan & Tijssen, 1990). Third, the results obtained from the full-count method are simpler to interpret (Okubo & Zitt, 2004).

3. The third step involved classification by scientific field. First, we classified the 7,155 journals in our sample according to the more than 200 categories listed in the WoS. Second, as WoS categories were too specific, we grouped them into 12 broad scientific disciplines using the Third European Report on S&T indicators.[3] In this classification, each WoS category is assigned to only one scientific discipline, but each journal is assigned to several categories by the WoS. If a journal is assigned to more than one scientific discipline, we again applied the full-count method so that we count one publication for each discipline.

3.2 Regional distribution of science across European regions

The spatial distribution of publications is mapped in Figure 1. As shown, of the 213 regions in total, 24 do not have any scientific publications, 34 have between one and 1,000 publications, 73 have between 1,001 and 6,000 publications, 43 have between 6,001 and 12,000 publications, 16 have between 12,001 and 18,000 publications and 23 have more than 18,000 publications.

[3] The classification was established by the Centre for Science and Technology Studies (CWTS) at Leiden University (see Tijssen and van Leeuwen, 2003). For categories not included in the CWTS 2003 classification, we used an updated (but unpublished) classification provided kindly by the CWTS.

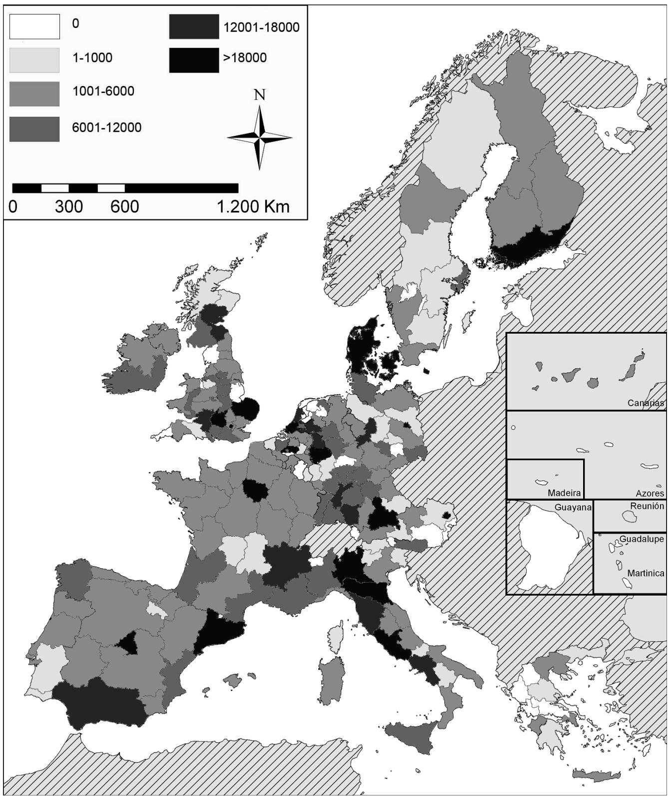


Figure 1. Spatial distribution of academic scientific papers by NUTS II in Europe 15 (1998–2004)

The indexes in Table 1 reveal that the production of scientific knowledge is highly concentrated in a few regions. As shown in Table 1, the Gini coefficient takes a value of 0.61 for the initial year (1998) and 0.59 for the latest year (2004) in the sample. Moreover, the trend—as shown in the Gini coefficients—is slightly downwards over the period 1998–2004. The remaining concentration indexes in Table 1 lead to the same conclusion; the C5 index takes a value of about 13, suggesting that just five regions account for 13% of papers. Similarly, the value of the C10 index is 22, indicating that 10 regions account for almost 22% of publications.

Table 1. Descriptive statistics and regional concentration indexes of academic scientific publications

	1998	1999	2000	2001	2002	2003	2004	98–04
N	157,446	164,492	166,660	170,603	174,266	179,770	193,398	1,206,644
Mean	739.19	772.27	782.49	800.96	818.16	844.00	907.98	5,664.99

Max.	5,794	5,950	5,887	6,162	6,186	6,401	6,701	43,081
Min.	0	0	0	0	0	0	0	0
Std. Dev.	937.41	972.52	976.27	995.60	1,013.52	1,046.23	1,100.44	7,024.09
C. Var.(1)	1.26	1.25	1.24	1.24	1.23	1.23	1.21	1.23
Coeff. Gini(2)	0.61	0.60	0.60	0.60	0.59	0.59	0.58	0.59
C5(3)	13.38	13.38	13.16	13.31	13.12	13.32	12.91	13.18
C10(4)	23.04	22.88	22.82	22.65	22.41	22.64	22.13	22.61
C25(5)	44.86	44.69	44.13	44.05	44.10	44.06	43.37	44.02

(1) Coefficient of variation = Std. Dev. ÷ Mean; (2) The Gini coefficient ranges between 0 and 1; the larger the value the higher the level of regional concentration in publications or collaborations. (3)(4)(5) Concentration indexes of publications for the top 5, 10 and 25 regions with the largest number of scientific papers, respectively.

In order to provide some descriptive details of how the production of scientific knowledge is distributed according to the level of regional development, we include in Table 2 information about the concentration of scientific production separating Objective 1 (regions where the GDP per capita is less than 75% of the European average) from the rest. Several facts emerge from this table:

1. The distinction between regions according to the level of economic development (GDP per capita) shows that less-developed NUTS regions generated 13.3% of all EU-15 academic papers in 1998. This percentage increased to 15.7% in 2004.

2. On average, Objective 1 regions produced 339 papers in 1998, while developed NUTS regions generated 904 papers in the same year. Therefore, the number of academic papers in a less-developed region was 37% of those generated in a developed NUTS region. This figure increased to 45% in 2004.

Several regions may be included in the group of developed regions, but have a low level of scientific capacity (e.g. those regions with a strong tourism sector). To present a complete picture, we also divide the regions according to the level of HERD per capita (right-hand side of Table 2). These data show that regions with less than 75% of the EU-15 average HERD per capita (42% of all regions in the sample) contributed to 12%

of all publications in 1998, increasing to 13.4% in 2004. On average, a region in this group produced 79% fewer papers than a region in the group with HERD per capita higher than 75% of the EU-15 average.

Table 2. Regional production of academic papers by type of NUTS region (*)

Groups of regions according to their level of development					Groups of regions according their level of HERD				
		1998	2004	1998–2004 % increase			1998	2004	1998–2004 % increase
Regions below 75% of the EU-15 average GDP per capita	No. Papers	20,996	30,463	45.09	Regions below 75% of the EU-15 average HERD per capita	No. Papers	12,064	15,855	31.42
	Mean	338.65	491.34			Mean	236.55	317.10	
	Std. Dev.	434.08	589.93			Std. Dev.	319.20	430.61	
Regions above 75% of the EU-15 average GDP per capita	No. Papers	136,450	162,935	19.41	Regions above 75% of the EU-15 average HERD per capita	No. Papers	86,905	102,274	17.68
	Mean	903.64	1,079.04			Mean	1,259.49	1,461.06	
	Std. Dev.	1,035.29	1,211.95			Std. Dev.	1,054.00	1,214.21	

(*) The group of less-developed regions comprises 62 NUTS regions, while the number of NUTS regions with more than 75% of the EU-15 average GDP per capita is 151 (213 in total). Because of the lack of data, the number of regions with less and more than 75% of the EU-15 average university R&D per capita falls to 120, with 51 NUTS regions in the first group and 69 in the second group.

This analysis shows an unbalanced picture of the generation of academic papers because the average capacity for publication of a less-developed region is about 45% of the capacity of a developed region in the core group. The disparities are rather stronger when we consider a classification of regions based on HERD expenditures.

Table 3 provides details of the 10 regions with the highest publications rate. Note that the UK accounts for three regions and Germany for two regions. The top 10 regions in terms of publications account for 22.61% of the total number of publications, which confirms that scientific knowledge production is highly concentrated in a few regions. Note that after data normalization, only the UK regions remain in the top 10 ranking.

Table 3. Regions with the highest number of academic publications and academic publications per capita (annual average 1998–2004)

	Annual average no. papers	%	Cum. (%)	Scientific publications/ population (in thousands)	
Île de France (FR10)	6,154	3.57	3.57	Inner London (UKI1)	3.01
Inner London (UKI1)	5,715	3.32	6.89	Prov. Vlaams-Brabant (BE24)	2.62
Denmark (DK00)	3,915	2.27	9.16	Berkshire, Buck. and Oxf. (UKJ1)	2.28
Oberbayern (DE21)	3,543	2.06	11.21	East Anglia (UKH1)	2.23
Lombardia (ITC4)	3,395	1.97	13.18	Wien (AT13)	1.85
Etelä-Suomi (FI18)	3,384	1.96	15.14	Groningen (NL11)	1.81
Berlin (DE30)	3,344	1.94	17.08	Eastern Scotland (UKM2)	1.76
East Anglia (UKH1)	3,296	1.91	19	Kärnten (AT21)	1.73
Berkshire, Buck. and Oxf. (UKJ1)	3,233	1.88	20.87	Gießen (DE72)	1.63
Cataluña (ES51)	2,989	1.73	22.61	Utrecht (NL31)	1.61
Others	133,409	77.39	100		
Annual average	172,378	100			

Following the procedure to retrieve the data explained above, we found that the top five scientific fields in terms of publications accounted for 67.94% of the total number of scientific publications in EU-15. These scientific fields are clinical medicine (17.33% of total number of publications), physics and astronomy (14.66%), chemistry (12.11%), biomedical sciences (11.98%) and basic life (11.86%). Table 4 lists the regions with the highest level of scientific production by discipline according to their share of total publications in each scientific field. Again, for every discipline the greatest numbers of publications are concentrated in just a few regions. The level of concentration of the 25 regions with more publications ranges from 40% to 50%. Note that Île de France appears as a top-five region in nine of the 12 scientific disciplines. It is also remarkable that Inner London (UK) appears as a top-five region across all of the scientific disciplines.

Table 4. Regions with the highest publications rate by scientific discipline and indexes of concentration (1998–2004)

Scientific field	Regions	No. of regional papers ÷ No. of papers (%)	Concentration indexes
	Denmark (DK00)	4.29	
1. Agricultural and food sciences	Gelderland (NL22)	3.99	C5 (16.32)
	Oberbayern (DE21)	2.85	C10 (27.65)
	Etelä-Suomi (FI18)	2.62	C25 (52.24)
	Inner London (UKI1)	2.57	
	Inner London (UKI1)	3.65	
2. Basic life	Île de France (FR10)	3.15	C5 (14.18)
	Denmark (DK00)	2.99	C10 (24.40)
	Etelä-Suomi (FI18)	2.2	C25 (46.68)
	East Anglia (UKH1)	2.19	
	Denmark (DK00)	3.59	
3. Biological sciences	Île de France (FR10)	2.8	C5 (15.94)
	Inner London (UKI1)	2.74	C10 (25.52)
	Gelderland (NL22)	2.39	C25 (45.29)
	Berk., Buc. and Oxf. (UKJ1)	2.26	
	Inner London (UKI1)	4.33	
4. Biomedical sciences	Lombardia (ITC4)	2.66	C5 (14.5)
	Berlin (DE30)	2.53	C10 (25.4)
	Denmark (DK00)	2.51	C25 (49.09)
	Oberbayern (DE21)	2.47	
	Île de France (FR10)	3.93	
5. Chemistry	Cataluña (ES51)	2.31	C5 (12.17)
	Inner London (UKI1)	2	C10 (20.89)
	Emilia-Romagna (ITD5)	1.97	C25 (40.95)
	Oberbayern (DE21)	1.96	
	Inner London (UKI1)	4.1	
6. Clinical medicine	Lombardia (ITC4)	3.39	C5 (15.75)
	Oberbayern (DE21)	2.91	C10 (27.77)
	Etelä-Suomi (FI18)	2.69	C25 (51.60)
	Zuid-Holland (NL33)	2.66	
	Île de France (FR10)	3.64	
7. Computer sciences	Inner London (UKI1)	3.27	C5 (12.99)
	Denmark (DK00)	2.13	C10 (21.98)
	Cataluña (ES51	2.03	C25 (42.56)
	Andalucía (ES61)	1.92	
	Île de France (FR10)	3.73	
8. Earth sciences	Denmark (DK00)	2.95	C5 (15.19)
	Inner London (UKI1)	2.93	C10 (24.65)
	East Anglia (UKH1)	2.8	C25 (43.58)
	Berk., Buc. and Oxf. (UKJ1)	2.78	
	Île de France (FR10)	3.58	
9. Engineering sciences	Inner London (UKI1)	3.44	C5 (12.95)
	Zuid-Holland (NL33)	2.17	C10 (21.34)
	East Anglia (UKH1)	1.95	C25 (41.1)
	Rhône-Alpes (FR71)	1.81	
	Île de France (FR10)	8.1	
10. Mathematics and statistics	Inner London (UKI1)	2.43	C5 (17.09)
	Andalucía (ES61)	2.35	C10 (25.84)
	Lazio (ITE4)	2.11	C25 (44.79)
	Comunidad de Madrid (ES30)	2.1	
	Île de France (FR10)	5.47	
11. Physics and astronomy	East Anglia (UKH1)	2.9	C5 (15.32)
	Inner London (UKI1)	2.82	C10 (24.80)
	Rhône-Alpes (FR71)	2.15	C25 (45.31)
	Lazio (ITE4)	1.98	
	Inner London (UKI1)	5.41	
12. Multidisciplinary	East Anglia (UKH1)	5.39	C5 (23.05)
	Berk., Buc. and Oxf. (UKJ1)	5.14	C10 (32.66)
	Île de France (FR10)	4.66	C25 (51.35)
	Denmark (DK00)	2.45	

(*) Concentration indexes of publications for the 5, 10 and 25 regions with the largest number of scientific papers. The numbering of disciplines shown here is used throughout subsequent tables.

3.3 Regional scientific specialization

The relative scientific specialization index is calculated in a similar way to the revealed technological advantage index (Soete and Wyatt, 1983):

$$\text{RSA}=\frac{P_{ij} / \sum_{j=1}^{12} P_{ij}}{\sum_{i=1}^{213} P_{ij} / \sum_{i=1}^{213}\sum_{j=1}^{12} P_{ij}} \qquad \text{Eq. 1}$$

where $P_{ij} / \sum_{j=1}^{12} P_{ij}$ is the number of publications of region *i* in discipline *j* over the number of publications of region *i* in all disciplines; $\sum_{i=1}^{213} P_{ij} / \sum_{i=1}^{213}\sum_{j=1}^{12} P_{ij}$ is the number of publications of all regions in discipline *j* over the total number of publications.

Table 5 shows the relative specialization indexes in each scientific field of the 20 regions with the highest number of papers. An index greater than one suggests a relative scientific strength of the region in that specific discipline. For example, those regions with the highest RSA are: Denmark[4] and Oberbayern in agriculture and food sciences; Noord-Holland and Köln in biomedical sciences; Cataluña and Emilia-Romagna in chemistry; Noord-Holland and Zuid-Holland in clinical medicine; and Comunidad de Madrid, Lazio and Cataluña in computer sciences.

---

[4] Note that in the case of Denmark, the NUTS II level is equal to the NUTS 0 level, i.e. the country level.

As the Herfindahl index shows, from this group of top science producers, Oberbayern, Emilia-Romagna and Karlsruhe have the highest concentration of publications by scientific discipline. Conversely, Zuid-Holland and Noord-Holland have the most diversified knowledge base.

Table 5. Relative scientific specialization of the 20 regions with the highest number of publications

	No. papers	1	2	3	4	5	6	7	8	9	10	11	12	H (*)
Île de France (FR10)	43,081	0.29	0.90	0.80	0.68	1.13	0.52	1.05	1.07	1.03	2.33	1.34	1.57	0.13
Inner London (UKI1)	40,003	0.78	1.11	0.83	1.31	0.61	1.25	0.99	0.89	1.05	0.74	1.64	0.85	0.12
Denmark (DK00)	27,402	1.84	1.28	1.54	1.08	0.69	0.97	0.91	1.26	0.75	0.73	1.05	0.76	0.15
Oberbayern (DE21)	24,804	1.42	1.04	0.71	1.23	0.98	1.45	0.63	0.60	0.54	0.51	1.02	0.92	0.17
Lombardia (ITC4)	23,766	0.75	1.03	0.56	1.36	0.73	1.73	0.83	0.56	0.64	0.77	0.52	0.79	0.14
Etelä-Suomi (FI18)	23,686	1.33	1.11	1.07	1.14	0.71	1.36	0.92	1.05	0.86	0.52	0.69	0.76	0.14
Berlin (DE30)	23,409	1.06	1.04	0.82	1.32	0.86	1.38	0.61	0.55	0.53	0.89	0.71	1.03	0.15
East Anglia (UKH1)	23,075	0.56	1.19	1.14	0.79	0.79	0.53	0.71	1.52	1.06	0.70	2.94	1.58	0.11
Berk., Buck. and Oxf. (UKJ1)	22,631	1.11	1.19	1.24	1.03	0.91	0.70	0.86	1.53	0.75	0.82	2.83	1.07	0.11
Cataluña (ES51)	20,923	1.25	1.10	1.11	0.91	1.30	0.82	1.14	1.06	0.87	1.15	0.65	0.88	0.11
Zuid-Holland (NL33)	20,267	0.24	1.08	0.64	1.24	0.66	1.56	1.12	0.59	1.27	0.69	0.98	0.72	0.10
Köln (DEA2)	19,861	0.58	0.97	0.87	1.28	0.78	1.45	0.68	0.66	0.51	0.80	0.77	1.19	0.13
Lazio (ITE4)	19,846	0.40	1.05	0.70	1.08	0.69	1.14	1.14	0.71	1.07	1.28	0.73	1.21	0.12
Wien (AT13)	18,999	1.51	0.93	0.98	1.16	0.76	1.44	0.97	0.78	0.69	0.78	0.64	0.86	0.13
Emilia-Romagna (ITD5)	18,767	0.92	0.95	0.71	1.12	1.26	1.13	0.91	0.71	0.87	0.74	0.50	1.02	0.16
Comunidad de Madrid (ES30)	18,394	1.32	1.07	1.05	1.03	1.07	0.62	1.16	0.71	1.00	1.38	0.78	1.19	0.13
Noord-Holland (NL32)	17,934	0.38	1.01	0.88	1.39	0.60	1.76	0.97	1.13	0.40	0.61	1.05	0.78	0.1
Prov. Vlaams-Brabant (BE24)	17,600	1.30	1.15	0.88	1.09	0.95	1.02	1.03	0.76	1.10	0.91	0.72	0.85	0.13
Karlsruhe (DE12)	16,953	0.14	0.81	0.37	1.09	0.96	1.52	1.02	0.89	0.85	0.99	0.88	1.09	0.16
Toscana (ITE1)	16,951	0.60	0.92	1.03	1.08	1.02	1.23	1.12	0.96	0.82	1.08	0.53	0.93	0.15

Notes: 1–12 are discipline identifiers as used in Table 4.
(*) Herfindahl concentration index of regional publications by scientific disciplines (the index takes a value of 1 when all the scientific papers published by a region occur in just one discipline, and 1/12 when the regional distribution of papers among scientific fields is the same across all disciplines).

3.4 Diagnosis of the scientific specialization patterns across regions

Following a methodology similar to Tuzi (2005), we assess regional scientific performance comparing the normalized revealed scientific advantage (RSA) to the normalized relative citation impact score (RCIS). The RSA is calculated as described in Section 3.3. The RCIS is obtained as the average number of citations per paper of region *i* in discipline *j* divided by the average number of citations per paper of all regions in discipline *j*. Then, both indexes are normalized into the range (–1,1); therefore, they equal 0 when the publications or citations of a region equal the EU-15 average.

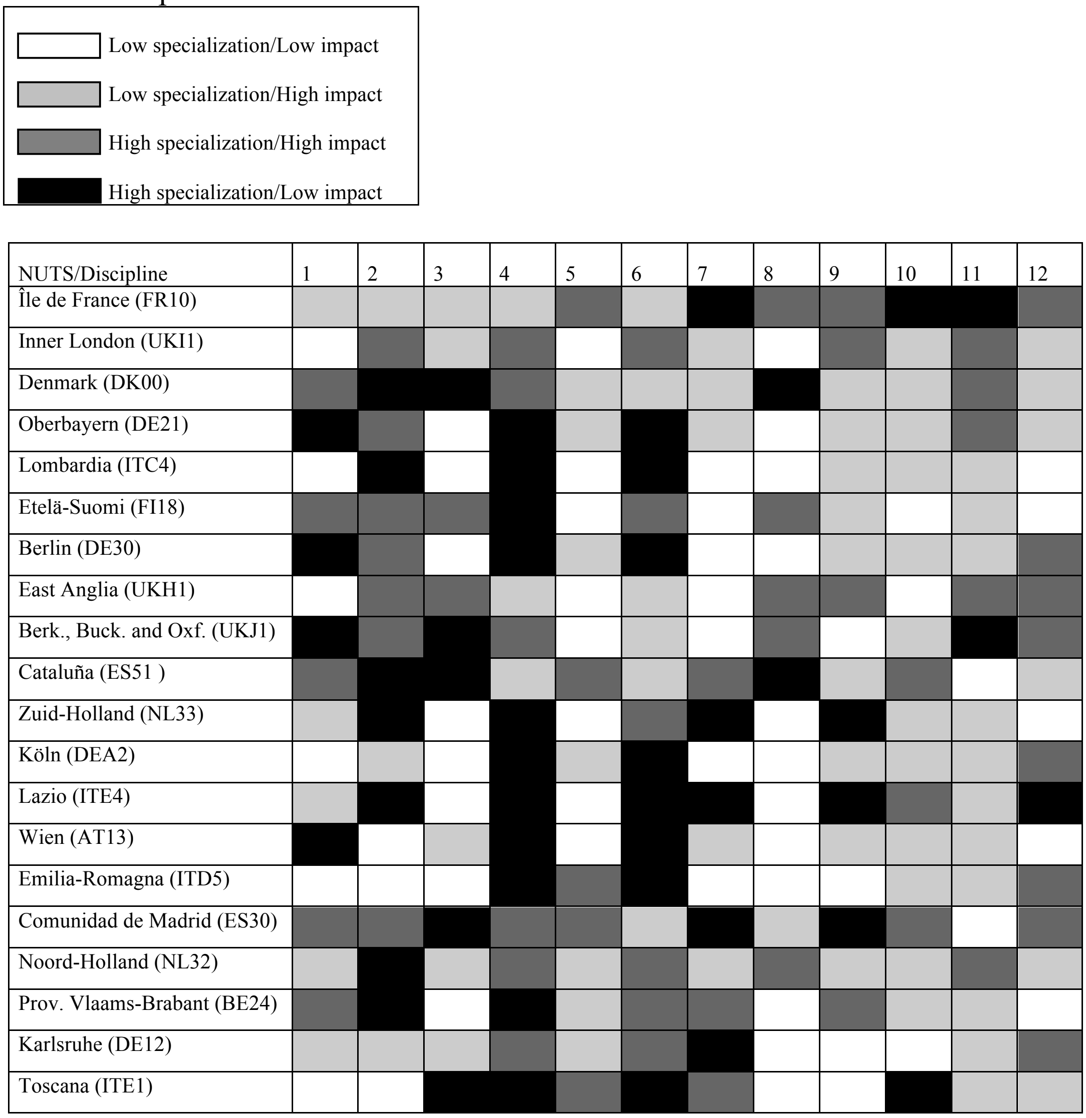
Table 6 proposes a classification of regions according to their scientific performance. On this basis, we identify four types of regions: a "superstar region" contributes a large share to EU-15 publications and citations in discipline *j*. On the contrary, a "capacity-lacking region" does not have a relative advantage in terms of quantity or quality of scientific production in discipline *j*. In between, there are regions that are performing high-quality but low-quantity research ("quality-focused regions), and vice versa ("quantity-focused regions").

Table 6. Types of regions according to their scientific performance

Impact (citations per paper)	–1 Low 0	0 High 1
High	**Quality oriented**	**Superstar**
Low	**Capacity lacking**	**Quantity oriented**
–1		

Regional relative specialization/Activity (publications)

Table 7. Evaluation of scientific specialization of the 20 regions with the largest number of publications



1. Agricultural and food sciences; 2. Basic life; 3. Biological sciences; 4. Biomedical sciences; 5. Chemistry; 6. Clinical medicine; 7. Computer sciences; 8. Earth sciences; 9. Engineering sciences; 10. Mathematics and statistics; 11. Physics and astronomy; 12. Multidisciplinary.

Following this classification, Table 7 shows the evaluation of the scientific specialization profiles of the 20 regions with the largest number of publications.

It should be noted that East Anglia and Comunidad de Madrid are superstars in six out of the 12 disciplines, and the former is not quantity oriented in any of them. On the contrary, Lazio is quantity focused in five scientific fields and a superstar in only one of them. Several opportunities for specialization arise in quality-oriented regions, such as Île de France, Denmark, Oberbayern, Köln, Wien and Noord-Holland. These results also show that biomedical sciences and clinical medicine are the disciplines in which most of the regions aimed at contributing to but focused on quantity rather than quality, because 11 and 9 regions, respectively, are quantity oriented.

4. Effects of HERD funding on regional scientific production

This section aims to address the questions related to the effects of academic R&D funding in promoting the production of scientific research at the regional level. The following subsections present the econometric model and the results.

4.1 Model and variables

We put forward a regional version of the KPF suggested by Adams and Griliches (1996) in terms of inputs and outputs. The inputs are academic R&D funds and the outputs are research publications. The empirical panel model takes the form:

$$\ln SP_{it} = \beta \ln RD(r)_{it} + \delta Sp(r)_{it} + \alpha_i + \eta_t + u_{it}$$

where the dependent variable SP_{it} is the scientific knowledge production of universities measured by the number of papers from region i in year t. The explanatory variables are as follows:

- ln*RD(r)* is the logarithm of past university R&D expenditures in the region (because it takes time for R&D to be reflected in new papers). This coefficient measures the returns to the scale of the regional research funds.*Sp* controls for the share that each scientific field has in the total production of scientific papers of the region. As we use aggregate data (production of papers), regions with a large participation in fields with a high propensity to publish are expected to produce more output.α represents regional-specific effects; it was included because research activity might be affected by several other contextual elements such as cultural practices, regional demand of research or particular regional scientific and innovation policies. This coefficient captures the changing level of regional efficiency to transform resources into results, if everything else were correctly specified in this equation.
- η captures time effects (for short panels, it is common to allow the time effect η_t to be a fixed effect including a set of time dummies).
- u is a disturbance term that captures all other unaccounted forces determining this particular measure of output.

It is worth noting that measurement problems arise when academic R&D expenditure is included as an explanatory factor. This variable is captured using R&D expenditure in the higher education sector (HERD) from Eurostat (in millions of purchasing power standard at 2000 prices). The variable includes all universities, colleges of technology, institutions of post-secondary education and centres operating under the control or associated to higher education institutions, whatever their source of funding or legal status. It should be kept in mind that on the one hand, academic R&D may underestimate the total value of the R&D resources of universities if there are other financial sources. On the other hand, R&D expenditures overstate the total value of resources devoted to academic scientific research because some of the R&D is assigned to the production of other outputs (for instance, university patents). Furthermore, the statistics also highlight the problem of missing observations for many European regions. Similar data restrictions to these have been stressed in the scarce research on this topic in other contexts; for example, Adams and Griliches (1996, 1998) reported comparable problems for the US and Crespi and Geuna (2008) pointed out related difficulties at the country level. We use this imperfect measure of academic R&D expenditure as we do not have a better input indicator.

As is known, the standard methods for the estimation of the suggested panel model are fixed effects or random effects. The major difference between these two techniques is the information for obtaining the coefficients. The fixed effects estimates are calculated from differences within each region across time; the random effects estimates are usually more efficient, because they include information across individual regions as well as across periods. The major drawback with random effects is that it is consistent only if the regional-specific effects are uncorrelated with the other explanatory variables. A Hausman specification test is frequently applied to determine whether the random effects estimation produces consistent estimates.

However, this standard procedure cannot be applied in this case for several reasons. First, we use an unbalanced panel in which HERD presents very little variance year by year. Second, the dependent variable (number of publications by regions) fluctuates over time (some regions have high values for one year and a dramatic drop for the following year). Given these characteristics of our sample, we first considered random effects estimation, and then applied additional well-known estimation procedures to determine the robustness of our results.

Finally, it is worth noting that spatial autocorrelation is not addressed in our model for theoretical reasons. The main channel for spillovers in research is through collaboration, where several factors such as specialization, sharing the same language or culture, or the economic distance can promote or hinder collaboration (Acosta et al., 2011b), but not the fact that they share a border. Below we discuss the empirical results in detail and perform additional econometric robustness checks.

4.2 Results

In order to analyse how the effect of HERD differs across regions according to their level of economic development, we estimate separate models for Objective 1 and non-Objective 1 regions. For each type of region, the models include HERD data lagged three and five years, respectively. The estimation strategy involves four models (Models I to IV in Table 8) where we explain the dependent variable (the number of scientific papers published by different regions and scientific fields, 1998–2004) with HERD, 12 scientific specialization variables and six dummy variables to capture year effects. As a referee pointed out, in order to obtain more accurate estimates of HERD it seems convenient to control for the different region sizes. Models V to VIII use the same estimation strategy but include population as a regressor.

All models are estimated using random effects estimators. The reason for using random effects is that, as explained above, HERD experiences little change year by year. The application of a within estimator (fixed effects) would produce misleading results because a differentiation of the variable HERD is required. The within estimates will be relatively imprecise for time-varying regressors (such as HERD) that vary little over time. In such cases, we are forced to use random effects estimation in order to learn anything about the population parameters (Wooldridge, 2002 p. 286). Nevertheless, below we apply alternative estimation procedures to check the reliability of our estimations.

Models I and II include the results for Objective 1 regions with 3 and 5 lags, respectively, for HERD; and Models III and IV present the estimations for non-Objective-1 regions. These models show significant coefficients for the variable HERD. Specialization and year effects also play a relevant role in explaining the production of scientific papers. Taking into account the size of the region produces lower values for the elasticities of HERD (Models V to VIII); however, note that the significance of HERD does not change in any of the models.

As indicated by previous research on this topic, the available data on university R&D prevent us from obtaining accurate estimates of elasticities. However, based on the results of the estimated baseline models presented in Table 8, we found some regularities that can be summarized as follows.

1. The coefficient of HERD is positive, significant and less than 1 in all models, suggesting both a positive and significant effect of R&D expenditure in the higher education sector (within the subsequent 3 and 5 years after the funding) and diminishing returns to scale.

2. The effects of HERD vary according to the level of development in regions. Estimates for the coefficient of university R&D are larger for Objective 1 regions than for developed regions. This means that an increase of 10%, for example, in university R&D expenditure has a larger impact (on average) on output (scientific papers) in Objective 1 regions than in developed regions. The explanation for this result might stem from several facts. First, the starting point in terms of production of scientific papers is higher in developed regions than in Objective 1 regions (note that elasticities give the impact in relative terms). Second, the specialization patterns and the weight of each field are different in the two types of regions. Third, developed regions might not have to rely as much on money as Objective 1 regions because the former counts on more experienced, better and more efficient scientific infrastructure, and in general more suitable conditions for research.

3. HERD take more time to produce scientific output in Objective 1 regions than in developed regions. In all our models in Table 8, we found that for Objective 1 regions, the elasticities take the largest value for the five-year lag; they increase from the third-year lag to the five-year lag. However, for developed regions, the elasticities more often take the largest value for the third lag and usually decrease to a minimum for a five-year lag. This finding can be explained because, as already said in the above paragraph,

developed regions rely on the scientific infrastructure in place, while non-Objective 1 regions need to build it.

4. Regional specialization has a significant effect for the production of scientific papers in regions; however, we obtained significant coefficients for fields in Objective 1 regions that are different from those obtained in more developed regions. This result suggests that scientific specialization matters for producing scientific output in both types of regions, but its effects seem to be strongly mediated by the specific scientific capacities of regions and other regional variables not specified in this model (such as the past accumulation of knowledge, the scientific related variety or the intensity of scientific collaborations in the region, etc.).

Table 8. Random effects estimates of HERD on the number of scientific papers published, by regions (1998–2004)

	Objective 1 regions		Non-Objective 1 regions		Objective 1 regions		Non-Objective 1 regions	
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
	t–3	t–5	t–3	t–5	t–3	t–5	t–3	t–5
HERD	0.5511*** (0.1408)	0.6020*** (0.0612)	0.1901*** (0.0466)	0.1271*** (0.0386)	0.3749** (0.1568)	0.4722*** (0.0483)	0.1246*** (0.0383)	0.0875*** (0.0332)
Pop					0.2954*** (0.1066)	0.2411*** (0.0510)	0.3842*** (0.0634)	0.4012*** (0.0640)
Sp1	0.6604 (1.4476)	0.9553 (0.8827)	–1.2013 (1.5612)	–0.6250 (1.5462)	0.8035 (1.3376)	0.9604 (0.8525)	–0.8342 (1.5302)	–0.3788 (1.5147)
Sp2	–2.2287 (2.2521)	–1.9814 (1.8981)	0.1041 (0.6444)	0.1643 (0.5872)	–2.0626 (2.1350)	–1.6520 (1.7556)	0.0738 (0.6465)	0.0980 (0.5947)
Sp3	–1.9249*** (0.3678)	–1.4409** (0.5725)	–0.4838** (0.2174)	–0.4544** (0.2194)	–1.6071*** (0.4225)	–1.1705** (0.4935)	–0.4611** (0.2177)	–0.4296* (0.2205)
Sp4	–0.7002 (1.5018)	–0.5502 (1.0672)	2.4734*** (0.8416)	2.0389** (0.7967)	–0.5647 (1.4413)	–0.4094 (0.9763)	2.2319*** (0.7927)	1.9104** (0.7671)
Sp5	0.3400 (1.0967)	–0.0314 (0.7738)	1.9532* (1.0358)	1.9403** (0.9688)	0.0571 (1.0818)	–0.0976 (0.7528)	1.9667** (0.9578)	1.9178** (0.9036)
Sp6	1.6845** (0.7131)	0.5976 (0.8885)	2.6422*** (0.5775)	2.9553*** (0.5820)	1.7541** (0.7231)	0.8822 (0.7917)	2.5716*** (0.5538)	2.7481*** (0.5430)
Sp7	0.0561 (1.1788)	1.2283 (0.7985)	0.7054 (0.4890)	0.6205 (0.4466)	0.0780 (1.1088)	1.0825 (0.7257)	0.6428 (0.4436)	0.5732 (0.4149)
Sp8	–1.9964 (2.1916)	–1.4103 (1.7444)	0.6395** (0.2845)	0.6544** (0.2856)	–1.8474 (2.0712)	–1.1551 (1.6551)	0.6000** (0.2591)	0.6168** (0.2585)
Sp9	–0.0627 (0.9575)	–0.0606 (0.5771)	1.1059 (0.7275)	1.3226* (0.6965)	0.0863 (0.8053)	0.0633 (0.5172)	1.1756* (0.7044)	1.3121** (0.6646)
Sp10	–0.7416 (1.2161)	1.0665 (1.1192)	–0.3515 (0.8275)	–0.2753 (0.7797)	–1.5137 (1.3255)	0.3221 (1.0282)	–0.3528 (0.8208)	–0.2822 (0.7819)
Sp11	–6.4121 (5.2307)	–4.2011 (5.2467)	2.7947* (1.6451)	3.0079** (1.5066)	–5.9044 (4.7828)	–3.9261 (4.6262)	2.3593 (1.5828)	2.5041* (1.4945)
Sp12	1.7112 (1.0421)	0.0953 (0.4959)	–0.1627 (0.4051)	–0.1944 (0.4268)	1.1848 (0.9876)	0.0966 (0.4658)	–0.2287 (0.3560)	–0.2454 (0.3717)
Y99	0.1012*** (0.0363)	0.1223*** (0.0413)	0.0106 (0.0232)	0.0165 (0.0206)	0.0882*** (0.0309)	0.1066*** (0.0338)	0.0107 (0.0210)	0.0143 (0.0199)

Y00	0.0814** (0.0394)	0.1250*** (0.0349)	–0.0670 (0.0413)	0.0610** (0.0255)	0.0865*** (0.0315)	0.1134*** (0.0274)	–0.0197 (0.0341)	0.0618** (0.0242)
Y01	0.1190*** (0.0424)	0.1560*** (0.0443)	–0.0797* (0.0459)	0.0422 (0.0326)	0.1336*** (0.0363)	0.1536*** (0.0360)	–0.0382 (0.0377)	0.0390 (0.0306)
Y02	0.1576*** (0.0384)	0.1836*** (0.0415)	–0.0694 (0.0455)	–0.0029 (0.0369)	0.1776*** (0.0351)	0.1875*** (0.0354)	–0.0293 (0.0375)	0.0087 (0.0334)
Y03	0.1943*** (0.0528)	0.1697*** (0.0336)	–0.0683 (0.0467)	0.0072 (0.0387)	0.2096*** (0.0465)	0.1783*** (0.0283)	–0.0314 (0.0375)	0.0122 (0.0344)
Y04	0.2854*** (0.0827)	0.1989*** (0.0466)	–0.0321 (0.0510)	0.0296 (0.0503)	0.2857*** (0.0698)	0.2136*** (0.0402)	–0.0005 (0.0430)	0.0327 (0.0460)
Cons	3.6759*** (0.9203)	3.7108*** (0.5327)	4.3103*** (0.6603)	4.4835*** (0.6044)	3.8049*** (0.8774)	3.5877*** (0.5315)	3.6708*** (0.6588)	3.7392*** (0.6212)
Wald chi^2	7996.9***	13032.9***	689.3***	1055.8***	7355.3***	25476.3***	735.5***	1073.3***
R^2	0.7619	0.8037	0.6956	0.6220	0.7467	0.7595	0.6621	0.6332
Spec. Eff.	1978.9 ***	1591.7***	444.9***	627.5***	1566.4***	2300.9***	424.3***	602.3***
Year Eff.	20.6***	37.4***	15.1**	13.7**	31.3***	63.1***	8.2	12.3*
No. obs.	143	143	425	425	143	143	425	425
No. regions	32	32	82	82	32	32	82	82

*p < 0.10, **p < 0.05, ***p < 0.01.

Notes: The dependent variable is the number of papers in logs. Robust standard errors are in parentheses. All regressions include time dummies. t–3 and t–5 are the lagged HERD variables for three and five years, respectively. R^2 is the overall R-squared for random effects. Sp1, Sp2, ..., Sp12 are regional specialization variables by discipline (number of scientific publications in each field divided by the total number of publications in the region). Year Effects and Specialization Effects are chi^2 tests (all year variables equal 0; all specialization variables equal 0).

4.3 Robustness checks

We check robustness in two ways. The first method is to compare the results with respect to the choice of econometric estimation method. To address the concern that there may still exist some unobserved or omitted variables across regions that drive the results in the baseline models presented in Table 8, we firstly consider fixed-effect estimation (FE). As it is known that this method allows the random term to be correlated with some exogenous regressors, it is used in empirical studies regularly. However, as explained before, there is little variance in HERD year by year, and therefore the fixed-effect estimation produces imprecise values. Consequently, it is not the best estimation procedure for studying the role of HERD in this particular case. We next consider the instrumental variables technique, which has proven to be useful when some variables have been omitted from the sampling model. Although a well-designed instrumental variables (IV) strategy is beyond the objective of this section because of the difficulty of finding an appropriate instrument, we have tried at least to analyse how the estimates and our main conclusion would change using, for example, regional GDP or regional population as an instrument. Table 9 includes FE and IV results for the main variable HERD (this table presents the results of IV using GDP as an instrument; using population produces quite similar coefficients).

Table 9. Alternative estimates of the effects of HERD on the number of scientific papers, by regions (1998–2004)

	Objective 1 regions		Non-Objective 1 regions	
	(I)	(II)	(III)	(IV)
	t–3	t–5	t–3	t–5
FE	0.2481 (0.1530)	0.5688*** (0.0585)	0.0650*** (0.0116)	0.0393** (0.0171)
No. obs.; No. regions	143; 32	143; 32	425; 82	425; 82
RE	0.3118** (0.1398)	0.5616*** (0.0551)	0.0854*** (0.0131)	0.0591*** (0.0154)
No. obs.; No. regions	143; 32	143; 32	425; 82	425; 82
IV (fe)	0.4011*** (0.0823)	0.8065*** (0.0780)	0.0961*** (0.0192)	0.0694*** (0.0197)
No. obs.; No. regions	115; 26	115; 26	365 ;70	365; 70
IV (re)	0.6732*** (0.0585)	0.7141*** (0.0480)	0.1238*** (0.0232)	0.1024*** (0.0235)
No. obs.; No. regions	115; 26	115; 26	365 ;70	365; 70

*p < 0.10, **p < 0.05, ***p < 0.01.

Notes: FE estimates control for the size of the region (using population), but year fixed effects are not included. RE is presented only for comparison purposes with FE (excluding year effects). IV uses the log of GDP in pps as an instrument and controls for specialization and year fixed effects. There are fewer observation for the IV estimates because of the lack of GDP data for some regions.

The second robustness check analyses the role of HERD using a three- or five-year lag, but this time applying a typical U-shaped function to R&D expenditures (Adams and Griliches, 1996; Crespi and Geuna, 2008). Table 10 presents the results using the same estimation procedures as before.

Table 10. Effects of HERD on the number of scientific papers with alternative distribution lags, by regions (1998–2004)

	Objective 1 regions		Non-Objective 1 regions	
	(I)	(II)	(III)	(IV)
	Weight 3t	Weight 5t	Weight 3t	Weight 5t
FE	0.6922^{***} (0.0982)	0.7813^{***} (0.0770)	0.1044^{***} (0.0207)	0.0854^{***} (0.0173)
No. obs.; No. regions	93; 23	93; 23	327; 82	327; 82
RE	0.6965^{***} (0.0927)	0.7841^{***} (0.0717)	0.1335^{***} (0.0226)	0.1038^{***} (0.0176)
No. obs.; No. regions	93; 23	93; 23	327; 82	327; 82
IV(fe)	0.8078^{***} (0.0560)	0.8955^{***} (0.0558)	0.1323^{***} (0.0222)	0.1048^{***} (0.0175)
No. obs.; No. regions	76; 17	76; 17	291; 70	291; 70
IV(re)	0.9118^{***} (0.0496)	0.9563^{***} (0.0501)	0.1640^{***} (0.0275)	0.1256^{***} (0.0222)
No. obs.; No. regions	76; 17	76; 17	291; 70	291; 70

*p < 0.10, **p < 0.05, ***p < 0.01.

Notes: Weight 3t: the inverted U-lag with weights 0.25, 0.5 and 0.25 for R&D lagged one, two and three years, respectively. Weight 5t: the inverted U-lag with weights 0.111, 0.222, 0.333, 0.222 and 0.111 for R&D lagged from one to five years, respectively. FE estimates control for region size (using population), but year fixed effects are not included. RE is presented only for comparison purposes with FE (excluding year effects). IV uses the log of GDP in pps as an instrument and controls for specialization and year fixed effects. There are fewer observations in this table than in Table 9 because we need data from consecutive years to obtain the weighted HERD. There are fewer observations for obtaining the IV estimates because of the lack of GDP data for some regions.

The results in both Table 9 and Table 10 show that despite changes in the values of the coefficients with respect to those in the baseline models (Table 8), our main conclusions still hold: the coefficients of HERD are always significant; the elasticities for the Objective 1 regions are higher than for non-Objective 1 regions; the values of the coefficients in the Objective 1 regions are higher for five-year lags than for three-year lags, and the opposite occurs for the non-Objective 1 regions.

5. Conclusions

This paper has attempted to identify the spatial distribution of academic scientific production across European regions, and it was mainly aimed at evaluating the role of HERD expenditures in encouraging academic scientific production. A preliminary descriptive analysis suggests a growing trend in the number of publications, increasing from 157,446 in 1998 to 193,398 in 2004. The data also display a high level of concentration of publications in a few regions, with little change over the period 1998–2004. For example, just five regions account for 13% of all publications, and this figure remained relatively unchanged over the period under examination.

The separation of regions according to different levels of economic development indicates that an Objective 1 region (one with a GDP per capita less than 75% of the EU-15 mean) produced on average less than half (45%) of the papers of a more economically advanced region. After dividing the NUTS regions into two groups according to HERD expenditure, the results show that a region in the less-favoured group (one with academic R&D per capita less than 75% of the EU-15 mean) produces on average 21% of the publications of a region in the group with R&D per capita expenditures greater than 75% of the EU-15 mean. Therefore, the descriptive statistics suggest that the level of development and the resources devoted to HERD affect the capacity to generate research outputs. From the evaluation of the scientific performance in each discipline of 20 regions with the largest number of publications, we found remarkable disparities between their relative contribution in terms of quantity and quality of their research.

In order to address the second group of research questions, related to the role of university R&D on regional scientific production, we estimated a KPF using random effect models. The base models were complemented with alternative estimates and lag structures for R&D expenditure. As in previous research on this topic, the available data on university R&D funds prevent us from obtaining accurate effects. Nevertheless, we have identified some regularities, as follows:

1. Money matters to produce scientific knowledge in universities across European regions, but there are decreasing returns to scale in the investment in HERD.

2. The effects are different according to the level of development in regions. Estimates for the coefficient of university R&D present larger values for Objective 1 regions than for developed regions. This might be the result of a lower starting point in Objective 1 regions than in developed regions, specific scientific specialization of these regions and lower dependency of developed regions than Objective 1 regions, because the former relies on more experienced, better and more efficient scientific infrastructure, and in general more suitable conditions for research.

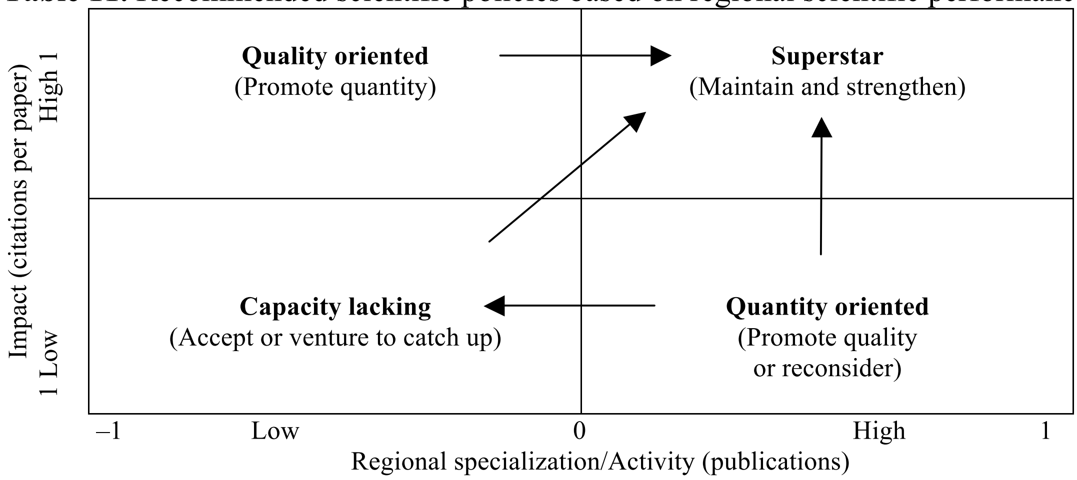
3. HERD expenditures take more time to produce scientific output in Objective 1 regions than in developed regions. This was confirmed using different lag structures and combinations.

4. Scientific specialization matters for producing scientific outputs in both types of regions, but its effects seem to be strongly mediated by the specific scientific capacities of regions and other regional variables not specified in this model (e.g. the past accumulation of knowledge, the scientific related variety or the intensity of scientific collaborations in the region, etc.).

Our results have some policy implications in the ERA framework, for cohesion policies, and in the context of a "smart scientific specialization". From a theoretical view, the rationality of university R&D public funds relies on correcting market failures arising from public goods, uncertainty and spillovers and enhancing the (non-linear) benefits of basic research, e.g. knowledge, trained people, equipment, problem solving, etc. When public intervention comes from a supranational government, rationality relies on seeking complementarities and coordination among policies. This paper has shown that HERD is a powerful tool to promote the quantity of regional scientific production. However, our results also show that their effects are time lagged, especially for Objective 1 regions. Policymakers should be aware of it when designing and implementing science and innovation policies for developing innovative capacities.

In the context of a smart specialization strategy, some policy implications emerge from the classification of regions according to their scientific performance. In *superstar regions*, science policy should be oriented to maintain and reinforce the existing scientific strength. In contrast, *capacity-lacking* regions could either focus on other disciplines in which they perform better or venture to develop the scientific capacity needed to become a superstar. Some opportunities for specialization may arise for those regions that are performing high-quality but low-quantity research ("quality-focused regions"). Finally, special attention is required for *quantity-focused* regions producing a large number of low-quality publications. The underlying causes of this situation should be detected. On the one hand, it may be the result of an incentive structure based on "publish all you can", no matter the quality. Then, the recommended strategy would be implementing policies oriented to promote research quality in order to become a superstar. On the other hand, the region may simply lack the scientific capacity to produce high-impact research, and thus two recommendations are suggested depending on the strategic priorities of these regions and their scientific potential. They could focus on other disciplines in which they are (or are likely to become) an EU-15 key player or they could work on the development of these capacities to catch up to superstar regions, being aware that they are falling behind. Obviously, all these science policies could have unintended consequences if there is no balance between the promotion of quantity and quality. For example, a quality-focused region that aims to increase its number of publications could end up publishing a large number of low-impact publications, and consequently losing its original competitive advantage.

Table 11. Recommended scientific policies based on regional scientific performance



Regarding the specific instruments to implement the above science policies, this paper showed that HERD are a powerful tool to promote regional scientific production in terms of quantity. Additionally, other instruments such as international scientific collaboration, attraction and retention of human capital, and mobility of researchers, are likely to contribute to the development of scientific capacity in terms of quantity and/or quality.

Overall, our results confirm that there is no sense in one-size-fits-all science and innovation policy, while it is strongly needed to introduce a tailor-made component in the regional science policy.

Finally, it is important to be aware of the limitations of the paper associated with the poor quality of the HERD data, with missing values for many of the regions and with little variance year by year, which causes some modelling problems. Additionally, we have not considered spillover effects in our models, although it is largely assumed that knowledge spills over across agents, regions and countries. This implies that a region benefits not only from its own investment in R&D but also from that of others. However, methodological difficulties in measuring and tracing spillovers hindered the authors from including them. Future research could aim to include spillover effects across regions and to address the effects of regional investment in R&D on the quality of the research.

Acknowledgements

The authors would like to thank to Robert Tijssen of the Centre for Science and Technology Studies (CWTS) at Leiden University for providing the updated classifications of scientific disciplines. Early versions of this paper were presented at the following conferences: 50th European Congress of the Regional Science Association International (ERSA 2010), 2011 ENID/STI Conference, and DRUID Society Conference 2011 and 2012. The authors would like to thank the conference participants for their helpful comments. The authors are grateful to Rodrigo Costas (CWTS) and Harro van Lente (Utrecht University) for their helpful insights in two seminars held at CWTS and Utrecht University in 2011. They also wish to thank two anonymous reviewers for their positive and constructive comments, which significantly contributed to improving the clarity and quality of the paper. Support for this research was provided in part from the Junta de Andalucía (Grant P08-SEJ-03981).

References

Abramovsky, L., Harrison, R., & Simpson, H. (2007). University Research and the Location of Business R&D. *The Economic Journal, 117*(519), pp. 114-141.

Acosta, M., & Coronado, D. (2003). Science-technology flows in Spanish regions: An analysis of scientific citations in patents. *Research Policy, 32*(10), pp. 1783-1803.

Acosta, M., Coronado, D., Ferrándiz, E., & León, M. (2011b). Factors affecting inter-regional academic scientific collaboration within Europe: the role of economic distance. *Scientometrics, 87*(1), pp. 63-74.

Acosta, M., Coronado, D., & Flores, E. (2011a). University spillovers and new business location in high-technology sectors: Spanish evidence. *Small Business Economics, 36*(3), pp. 365-376.

Adams, J. D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981-1999. *Research Policy, 34*(3), pp. 259-285.

Adams, J. D., & Griliches, Z. (1996). Research Productivity in a System of Universities. *NBER Working Paper Series 5833*.

Adams, J. D., & Griliches, Z. (1998). Research Productivity in a System of Universities. *Annales d'Economie et de Statistique*(49-50), pp. 127-162.

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., & Sapir, A. (2007). Why Reform Europe's Universities? *Bruegel Policy Brief, 4*, pp. 1-8.

Agrawal, A., & Cockburn, I. (2003). The anchor tenant hypothesis: exploring the role of large, local, R&D-intensive firms in regional innovation systems. *International Journal of Industrial Organization, 21*(9), pp. 1227-1253.

Anselin, L., Varga, A., & Acs, Z. (1997). Local geographic spillovers between university research and high technology innovations. *Journal of Urban Economics, 42*(3), pp. 422-448.

Anselin, L., Varga, A., & Acs, Z. J. (2000). Geographic and sectoral characteristics of academic knowledge externalities. *Papers in Regional Science, 79*(4), pp. 435-443.

Arora, A., David, P. A., & Gambardella, A. (1998). Reputation and Competence in Publicly Funded Science: Estimating the Effects on Research Group Productivity. *Annales d'Economie et de Statistique*(49-50).

Audretsch, D. B., Lehmann, E. E., & Warning, S. (2005). University spillovers and new firm location. *Research Policy, 34*(7), pp. 1113-1122.

Beise, M., & Stahl, H. (1999). Public research and industrial innovations in Germany. *Research Policy, 28*(4), pp. 397-422.

Bordons, M., Fernández, M., & Gómez, I. (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance. *Scientometrics, 53*(2), pp. 195-206.

Calderini, M., & Scellato, G. (2005). Academic research, technological specialization and the innovation performance in European regions: an empirical analysis in the wireless sector. *Industrial and Corporate Change, 14*(2), pp. 279-305.

Carayol, N., & Matt, M. (2006). Individual and collective determinants of academic scientists' productivity. [doi: 10.1016/j.infoecopol.2005.09.002]. *Information Economics and Policy, 18*(1), pp. 55-72.

Cesaroni, F., Gambardella, A., (2003) Research productivity and the allocation of resources in publicly funded research programmes, in: A. Geuna, A. Salter & W. E. Steinmueller (Eds.) *Science and Innovation: Rethinking the Rationales for Funding and Governance* (Cheltenham: Edward Elgar).

Cohen, W. M., Florida, R., Randazzese, L., & Walsh, J. P. (1998) Industry and the Academy: Unease partners in the cause of technological advance in: R. G. Noll (Ed.) *Challenges to Research Universities* (Washington, DC.: Brookings Institute Press).

Crespi, G. A., & Geuna, A. (2008). An empirical study of scientific production: A cross country analysis, 1981-2002. [doi: DOI: 10.1016/j.respol.2007.12.007]. *Research Policy, 37*(4), pp. 565-579.

Dasgupta, P., & David, P. (1994). Towards a new economics of science. *Research Policy, 23*, pp. 487-521.

de Dominicis, L., Elena Pérez, S., & Fernández-Zubieta, A. (2011). European university funding and financial autonomy. A study on the degree of diversification of university budget and the share of competitive funding. *Scientific and Technical Research series*. Luxembourg: Publications Office of the European Union.

European Commission. (2000). Presidency Conclusions of the Lisbon European Council. Brussels: Commission of the European Communities.

European Commission. (2001). The Regional Dimension of the European Research Area. Brussels: Commission of the European Communities.

European Commission. (2007). Green paper "The European Research Area. New perspectives". Brussels: Commission of the European Communities.

European Commission. (2010). Regional Policy contributing to smart growth in Europe 2020. Brussels: Commission of the European Communities.

Fischer, M., & Varga, A. (2003). Spatial knowledge spillovers and university research: Evidence from Austria. *Annals of Regional Science, 37*(2), pp. 303-322.

Gauffriau, M., Larsen, P., Maye, I., Roulin-Perriard, A., & von Ins, M. (2008). Comparisons of results of publication counting using different methods. *Scientometrics, 77*(1), pp. 147-176.

Hoekman, J., Frenken, K., & Oort, F. (2009). The geography of collaborative knowledge production in Europe. [doi:10.1007/s00168-008-0252-9]. *Annals of Regional Science, 43*, pp. 721-738.

Jaffe, A. (1989). Real effects of academic research. *American Economic Review, 79*(5), pp. 957-970.

Jaffe, A., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics, 79*(3), pp. 577-598.

Mansfield, E. (1991). Academic research and innovation. *Research Policy, 20*, pp. 1-12.

Mansfield, E. (1998). Academic research and industrial innovation: An update of empirical findings. *Research Policy, 26*(7-8), pp. 773-776.

Mansfield, E., & Lee, J.-Y. (1996). The modern university: contributor to industrial innovation and recipient of industrial R&D support. *Research Policy, 25*(7), pp. 1047-1058.

Maurseth, P. B., & Verspagen, B. (2002). Knowledge Spillovers in Europe: A Patent Citations Analysis. *Scandinavian Journal of Economics, 104*(4), pp. 531-545.

Okubo, Y., & Zitt, M. (2004). Searching for research integration across Europe: a closer look at international and inter-regional collaboration in France. [doi:10.3152/147154304781780019]. *Science and Public Policy, 31*, pp. 213-226.

Payne, A., & Siow, A. (2003). Does Federal Research Funding Increase University Research Output? *Advances in Economic Analysis & Policy, 3*(1), p. 1018.

Soete, L., & Wyatt, S. (1983). The use of foreign patenting as an internationally comparable science and technology output indicator. *Scientometrics, 5*(1), pp. 31-54.

Stephan, P. E. (1996). The Economics of Science. *Journal of Economic Literature, 34*(3), pp. 1199-1235.

Tijssen, R., J.W, & van Leeuwen, T. N. (2003). Bibliometric Analyses of World Science, Extended Technical Annex to Chapter 5 of the Third European Report on Science & Technology Indicators.

Tuzi, F. (2005). The scientific specialisation of the Italian regions. *Scientometrics, 62*(1), pp. 87-111.

van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics, 62*(1), pp. 133-143.

van Raan, A. F. J., & Tijssen, R. J. W. (1990) An overview of quantitative science and technology indicators based on bibliometric methods *Technology Economy Programme for the development of Indicators.*Paris: OECD. Cited from Gauffriau et al,(2008).

Verspagen, B., & Schoenmakers, W. (2000). The Spatial Dimension of Knowledge Spillovers in Europe: Evidence from Firm Patenting Data

Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics, 62*(1), pp. 117-131.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.